CORE
Metadata, citation and similar papers at core.ac.uk
Provided by UKM Journal Article Repository

Sains Malaysiana 48(5)(2019): 1151–1156
http://dx.doi.org/10.17576/jsm-2019-4805-24

A Comparison of Asymptotic and Bootstrapping Approach in Constructing Confidence Interval of the Concentration Parameter in von Mises Distribution

(Perbandingan Pendekatan Asimptot dan Pembutstrapan dalam Membina Selang Keyakinan Parameter Menumpu bagi Taburan von Mises)

NOR HAFIZAH MOSLIM, YONG ZULINA ZUBAIRI*, ABDUL GHAPOR HUSSIN,
SITI FATIMAH HASSAN & NURKHAIRANY AMYRA MOKHTAR

ABSTRACT

Bootstrap is a resampling procedure for estimating the distributions of statistics based on independent observations. Basically, bootstrapping has been established for the use of parameter estimation of linear data. Thus, the used of bootstrap in confidence interval of the concentration parameter, κ in von Mises distribution which fitted the circular data is discussed in this paper. The von Mises distribution is the 'natural' analogue on the circle of the Normal distribution on the real line and widely used to describe circular variables. The distribution has two parameters, namely mean direction, μ and concentration parameter, κ, respectively. The confidence interval based on the calibration bootstrap method will be compared with the existing method, confidence interval based on the asymptotic to the distribution of $\hat{\kappa}$*. Simulation studies were conducted to examine the empirical performance of the confidence intervals. Numerical results suggest the superiority of the proposed method based on measures of coverage probability and expected length. The confidence intervals were illustrated using daily wind direction data recorded at maximum wind speed for seven stations in Malaysia. From point estimates of the concentration parameter and the respective confidence interval, we note that the method works well for a wide range of κ values. This study suggests that the method of obtaining the confidence intervals can be applied with ease and provides good estimates.*

Keywords: Calibration bootstrap; circular variable; concentration parameter; von Mises distribution

ABSTRAK

Kaedah pembutstrapan adalah proses persampelan semula data bagi menganggarkan taburan statistik berdasarkan pemerhatian bebas. Kebelakangan ini, kaedah pembutstrapan telah digunakan secara meluas untuk menganggar parameter data linear. Oleh itu, dalam kajian ini, kami menggunakan kaedah pembutstrapan dalam membina selang keyakinan terhadap parameter menumpu, κ bagi taburan von Mises. Taburan von Mises dikenali sebagai taburan normal membulat dan ia merupakan taburan yang menyerupai taburan normal seperti yang biasa digunakan dalam statistik linear. Taburan ini mempunyai dua parameter, iaitu min berarah, μ dan parameter menumpu, κ. Selang keyakinan berdasarkan kaedah pembutstrapan penentukuran akan dibandingkan dengan kaedah sedia ada, selang keyakinan berdasarkan asimptotik $\hat{\kappa}$*. Kajian simulasi dan penilaian bagi saiz selang dan kebarangkalian menumpu telah dijalankan bagi menilai ketepatan empirik selang keyakinan tersebut. Kaedah ini diilustrasikan menggunakan data arah angin harian yang dirakamkan pada kelajuan angin maksimum bagi tujuh stesen di Malaysia. Titik penganggaran bagi parameter menumpu dan selang keyakinan, masing-masing menunjukkan kaedah pembutstrapan penentukuran ini berfungsi dengan baik untuk pelbagai nilai κ. Kajian ini menunjukkan bahawa kaedah mendapatkan selang keyakinan boleh digunakan dengan mudah dan memberikan anggaran yang baik.*

Kata kunci: Parameter menumpu; pemboleh ubah membulat; pembutstrapan penentukuran; taburan von Mises

INTRODUCTION

Statistical data can be classified according to their distributional topologies. A linear data set can be represented on a straight line and for circular data, they can be represented by the circumference of a unit circle. For circular data, they are commonly measured in the range of [0°, 360°) degrees or [0, 2π) radian. It is worthwhile to note that statistical theories for straight line and circle are very different from one to another because the circle is a closed curve. Circular or directional data can be found in the area of meteorology such as wind direction. Additionally, such data can be found in other fields such as in movement coordination (Stock et al. 2018), solar radiation (Polo et al. 2018), biology and physics (Fitak et al. 2018), hydrology (Yan et al. 2016) and remote sensing (Kucwaj et al. 2017).

Commonly used distribution to describe circular random variable is the von Mises distribution. The density function is given as

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left(\kappa\cos(\theta-\mu)\right), \quad 0 \le \theta < 2\pi \qquad (1)$$

where $0 \le \mu < 2\pi$ and $0 < \kappa < \infty$ are the parameters. $I_0(\kappa)$ is the modified Bessel function of order zero and can be defined as,

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \exp\left(\kappa \cos\left(\theta - \mu\right)\right) d\theta \tag{2}$$

Approximate solution of the concentration parameter, $\hat{\kappa}$ can be obtained by the maximum likelihood estimation (Best & Fisher 1981) which was defined as follows,

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + \frac{5}{6}\bar{R}^5 & \text{if } \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + \frac{0.43}{1-\bar{R}} & \text{, if } 0.53 \le \bar{R} < 0.85 \\ \frac{1}{\left(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}\right)} & \text{if } \bar{R} \ge 0.85 \end{cases} \tag{3}$$

where $\bar{R} = \sqrt{\left(\bar{C}^2 + \bar{S}^2\right)}$, $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} \cos\theta_i$ and $\bar{S} = \frac{1}{n}\sum_{i=1}^{n} \sin\theta_i$.

This is a continuous probability distribution and as κ approaches 0, the distribution converges to the uniform distribution. Meanwhile, as κ increase the distribution converges to the point distribution concentrated in the direction μ. Thus, it will approach the normal distribution with the mean μ_0 and variance $\frac{1}{\kappa}$ (Fisher 1993; Mardia 1972). Since then, von Mises distribution can also be called as Circular Normal Distribution as it has the similarities with the normal distribution on the real line (Fisher 1993). The approximation of concentration parameter, κ to the normal distribution has been tested earlier through the simulation study (Moslim et al. 2017).

In data analysis, confidence interval is often used as they combine both point estimate and hypothesis testing into a single inferential statement. In other words, confidence interval gives an estimated range of values which is likely to include an unknown population parameter with a specified probability within that interval. A number of studies were done to approximate confidence interval for the concentration parameter of von Mises distribution including those using bootstrap (Hassan et al. 2014; Khanabsakdi 1996; Stephens 1969). However, some methods may not have good coverage accuracy of the confidence interval if size of data is small (Fisher 1993). In this paper, we propose calibration bootstrap for the concentration parameter of von Mises distribution. It is assumed that the method will have an improvement on the coverage probability and robust against data distribution. Simulation studies will be carried out for the proposed method and compared with the standard approach using asymptotic distribution of $\hat{\kappa}$. Their performance is examined using coverage probability and expected length values (Hassan et al. 2014, 2012; Letson & McCullough 1998).

ASYMPTOTIC DISTRIBUTION OF $\hat{\kappa}$

Confidence interval based on the distribution of $\hat{\kappa}$ is normally distributed with mean and variance as follows (Jammalamadaka & SenGupta 2001):

$$\hat{\kappa} \sim N\left(\kappa, \frac{1}{n\left(1 - \frac{\bar{R}}{\hat{\kappa}} - \bar{R}^2\right)}\right) \tag{4}$$

Then, the 95% confidence interval for the concentration parameter,κ can be found as follows:

$$\left(-B + \hat{\kappa}_1, B + \hat{\kappa}_1\right) \tag{5}$$

where

$$B = \frac{1.96}{\left[n\left(1 - \frac{\bar{R}}{\hat{\kappa}} - \bar{R}^2\right)\right]^{1/2}},\ \bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2},\ \bar{C} = \frac{1}{n}\sum_{i=1}^{n}\cos\theta_i \text{ and}$$

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\sin\theta_i.$$

CALIBRATION BOOTSTRAP

Bootstrap method is a computer-based technique for making certain kind of statistical inferences which can simplify the often-intricate calculations of traditional statistical theory (Efron 1979). This method was introduced as a nonparametric device for estimating standard errors and biases.

One early method of obtaining confidence interval for concentration parameter is using the percentile bootstrap method (Fisher 1993). This approach is further improved using bootstrap-*t* (Hassan et al. 2014). However, the bootstrap-*t* method only limits to the second order accuracy and the algorithm can be numerically unstable. Thus, to ensure good coverage accuracy and overall expected length, we propose a calibration bootstrap method which improves to the third-order accuracy (DiCiccio & Efron 1996).

Calibration is a bootstrap resampling technique that performs a second bootstrap loop. Let

$$p\left(\hat{\lambda}\right) = \mathbf{Prob}\left\{\theta \le \hat{\theta}_{\hat{\lambda}}\right\} = \alpha \tag{6}$$

Once the value of $\hat{\lambda}$ is obtained and if the procedure is calibrated correctly, the value of $\lambda = \alpha$ is achieved. Let

$$\hat{p}\left(\lambda\right) = \mathbf{Prob}_*\left\{\hat{\theta} \le \hat{\theta}^*_{\lambda}\right\} \tag{7}$$

the bootstrap estimate of $p(\lambda)$ where "*" refers to the bootstrap sampling and $\hat{\theta}$ is fixed. Generate a number of bootstrap samples then compute $\hat{\theta}^*_{\lambda}$ for each one and record the proportion of times that $\hat{\theta} \le \hat{\theta}^*_{\lambda}$. By using the same bootstrap samples, the process is repeated for a range of

λ values that includes the nominal value α. The value of λ that satisfy $\hat{p}(\lambda) = \alpha$ is denoted by $\hat{\lambda}_\alpha$.

The following steps describe the calibration bootstrap method:

1. Generate n values of θ_i from the $VM(\mu,\kappa)$ where $0 \leq \theta < 2\pi$ and $i = 1, 2, \ldots, n$.
2. Estimate the bootstrap parameter for the bootstrap samples from (1) and label it as $\hat{\kappa}_1$.
3. Repeat step (1) and (2) to obtain B bootstrap parameter estimates; $\hat{\kappa}_b$ where $b = 1, 2, K, B$.
4. For each bootstrap samples, compute a λ-level confidence point $\hat{\theta}^*_\lambda(b)$ for a range of values of λ.
5. Get the value of $\hat{p}(\lambda) = \#\left\{\hat{\theta} \leq \hat{\theta}^*_\lambda(b)\right\} / B$ for each λ.
6. Find the value of λ that satisfy $p(\lambda) = \alpha$.

SIMULATION STUDY

Simulation study were conducted for three different sample sizes, n = 30, 50 and 100 with four values of concentration parameter, κ = 2, 4, 6 and 8. Without loss of generality, the mean direction value will be assumed as 0. The significance level for the asymptotic distribution of $\hat{\kappa}$ is set at $\alpha = 0.05$ meanwhile for the calibration bootstrap method is $\alpha = 0.01$. This has been evaluated previously to be the suitable value of α to get the probability of 0.95. The number of bootstrap replications, B for each simulation is set at 100 (Efron & Tibshirani 1993).

Let s be the number of simulation studies and it was repeated for 1020 times. Two indicators to determine the best method in constructing intervals were calculated as follows:

coverage probability $= \frac{m}{s}$, where m is the number of true value that fall within the confidence interval and expected length = upper limit - lower limit.

Coverage probability is the proportion number that the confidence interval contains the true value of concentration parameter for each method. The confidence level considered in this study is 95%. Thus, the best result is measured through the coverage probability value that is close to 0.95. Expected length is the class size of a confidence interval. It is another indicator to determine the best method of constructing the confidence interval. The best and efficient method will give the shortest expected length.

RESULTS AND DISCUSSION

Tables 1 and 2 show the coverage probability and expected length for all values of concentration parameter, κ and sample sizes, n for each method, respectively. Each method is labelled as follows:

M1 - asymptotic distribution of $\hat{\kappa}$, and
M2 - calibration bootstrap

From the results as displayed in Table 1, with the exception of n = 30, as we increase the values of $\hat{\kappa}$, the coverage probability using M2 outperforms the M1 method.

From the result in Table 2, the expected length of the confidence interval for both methods decreases as the sample size increases. It is worthwhile to note that the expected length increases as the concentration parameter increases. With the exception of n = 30, the expected length of M1 and M2 are very close to each other with M1 consistently smaller.

Based on the two measures, we note that the two methods are comparable. However, coverage probability is often used to measure performance of the confidence interval (Letson & McCullough 1998. In this case, based on coverage probability, M2 is the superior method or in

TABLE 1. Coverage probability for concentration parameter, $\kappa = 2, 4, 6, 8$ and sample sizes, $n = 30, 50, 100$

n		30		50		100	
Method		M1	M2	M1	M2	M1	M2
	2	0.97	0.93	0.97	0.94	0.96	0.95
	4	0.96	0.92	0.98	0.94	0.94	0.95
κ	6	0.97	0.92	0.96	0.94	0.94	0.95
	8	0.96	0.92	0.95	0.94	0.95	0.95

TABLE 2. Expected length for concentration parameter, $\kappa = 2, 4, 6, 8$ and sample sizes, $n = 30, 50, 100$

n		30		50		100	
Method		M1	M2	M1	M2	M1	M2
	2	1.9327	2.5214	1.4333	1.7356	0.9903	1.1369
	4	4.1183	5.5207	3.0595	3.7399	2.1200	2.4098
κ	6	6.1525	8.3145	4.7402	5.8648	3.2872	3.7858
	8	8.5748	11.5521	6.4970	7.9670	4.3865	5.1279

other words, calibration bootstrap is a superior method.

As mentioned earlier, calibration bootstrap could provide good approximation for confidence interval. Although calibration bootstrap is computationally intensive, this can be easily overcome with the advancement of technology and supercomputing facilities (Lv et al. 2017). Most importantly, the convergence rate is faster compared to other uncalibrated method (Loh 1991). Furthermore, calibration bootstrap is distribution-free and thus adds to the simplicity of method (DiCiccio & Efron 1996). In contrast, for confidence interval based on the asymptotic distribution of $\hat{\kappa}$, one needs to check on the assumption of normality.

ILLUSTRATIVE EXAMPLE

As an illustration of the proposed method, daily wind direction data (in radian) recorded at maximum wind speed (in m/s) for seven stations in Malaysia were considered. These data were collected at the year of 2008 at an altitude of 2.3 m to 37.8 m. A total of 70 data points was obtained from each station during the northeast monsoon (November to March). For the peninsular Malaysia, the stations are located at west coast and east coast regions. For the west coast region, the stations are located at Alor Setar, Kuala Lumpur International Airport (KLIA), Melaka and Senai meanwhile for the east coast region, the stations are located at Kuala Terengganu and Kota Bharu as shown in Figure 1. One station located at east Malaysia which is Kota Kinabalu was considered as well. All data were obtained from Malaysian Meteorological Department.

Satari et al. (2015) analysed the characteristics of Malaysian wind direction that was recorded at maximum wind speed for the years of 1999 to 2008. Those data described nicely with the von Mises distribution. Thus, we test the suitability of all stations to the von Mises distribution by plotting the goodness of fit based on quartiles using Oriana software. Due to the limited space, Figure 2 shows the goodness of fit plot to the von Mises distribution based on quantiles for the wind direction data at Kuala Terengganu station. From the figure, it can be seen that the wind direction data of all seven stations fits well to the von Mises distribution.

Table 3 shows the upper limit, lower limit and expected length for the calibration bootstrap method from the stations located at west coast region, east coast region and east Malaysia. For each station, the concentration parameter, $\hat{\kappa}$ was calculated. It can be seen that, the concentration parameter value for the east coast region was higher than the west coast region. In the west coast region, Melaka recorded the highest concentration parameter than other stations. Similarly, Kuala Terengganu recorded the highest concentration parameter than other stations at east coast region. Thus, the wind direction at both stations is less scattered and less dispersed. It is worthwhile to note that, Kota Kinabalu had the lowest concentration parameter suggesting that the wind direction at east Malaysia was more scattered and more dispersed during the northeast monsoon.

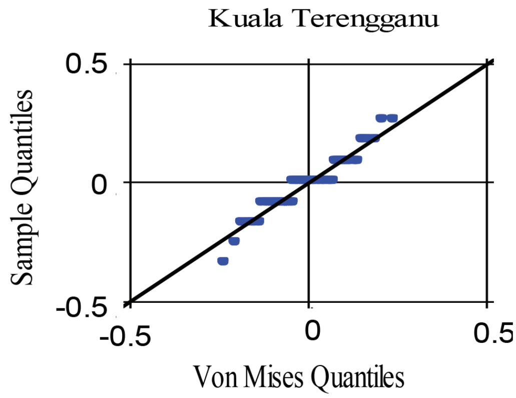

FIGURE 2. Goodness of fit plot to the von Mises distribution of wind direction data

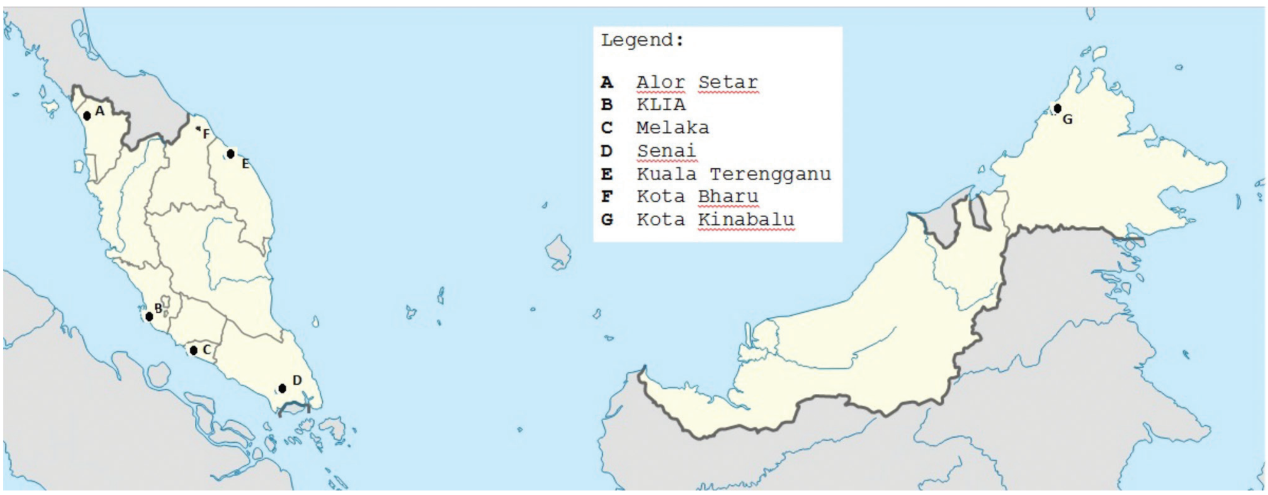

FIGURE 1. Malaysia Map (source: ms.wikipedia.org/wiki/Fail:Malaysia_location_map.svg)

TABLE 3. Confidence Interval (CI) for wind direction data recorded at maximum wind speed

Region	Station	$\hat{\kappa}$	Confidence interval		Expected length	
			M1	M2	M1	M2
West Coast, Peninsular Malaysia	Alor Setar	0.9533	(0.5644,1.3421)	(0.6532,1.4959)	0.7777	0.8427
	KLIA	0.9879	(0.5946,1.3812)	(0.5810,1.4967)	0.7866	0.9157
	Melaka	1.7617	(1.2344,2.2890)	(1.3272,2.3907)	1.0547	1.0635
	Senai	1.5844	(1.0939,2.0748)	(1.1853,2.2280)	0.9809	1.0427
East Coast, Peninsular Malaysia	Kuala Terengganu	9.0689	(6.1523,11.9856)	(7.4815,13.6182)	5.8333	6.1367
	Kota Bharu	5.1690	(3.5490,6.7889)	(3.7628,7.4614)	3.2399	3.6986
East Malaysia	Kota Kinabalu	0.8214	(0.4477,1.1950)	(0.4879,1.2922)	0.7473	0.8042

CONCLUSION

This article proposed calibration bootstrap method in constructing the confidence interval of the concentration parameter for the von Mises distribution. The method is derived and compared to the asymptotic distribution of $\hat{\kappa}$. The performances were evaluated via simulation study for various values of sample size and concentration parameter. The coverage probability value is more influenced in measuring the performance of confidence interval. Consequently, calibration bootstrap performs better result compared to the other method. The methods were tested using real data set and the results aligned with the simulation results.

ACKNOWLEDGEMENTS

The authors wish to thank University of Malaya (BK045-2016) and (GPF006H-2018) for the research grant in undertaking this research.

REFERENCES

Best, D.J. & Fisher, N.I. 1981. *The BIAS of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. Communication in Statistics-Simulation and Computation* 10(5): 493-502.

DiCiccio, T.J. & Efron, B. 1996. *Bootstrap confidence interval. Statistical Science* 11(3): 189-228.

Efron, B. & Tibshirani, R.J. 1993. *An Introduction to the Bootstrap.* London: Chapman & Hall.

Efron, B. 1979. *Bootstrap methods: Another look at the jackknife. The Annals of Statistics* **7**(1): 1-26.

Fisher, N. 1993. *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.

Fitak, R.R., Caves, E.M. & Johnsen, S. 2018. Orientation in pill bugs: An interdisciplinary activity to engage students in concepts of biology, physics & circular statistics. *American Biology Teacher* 80(8): 608-618.

Hassan, S.F., Hussin, A.G. & Zubairi, Y.Z. 2012. *Improved efficient approximation of concentration parameter and confidence interval for circular distribution. ScienceAsia* 38(1): 118-124.

Hassan, S.F., Zubairi, Y.Z., Hussin, A.G. & Satari, S.Z. 2014. *Some confidence intervals for large concentration parameter in von Mises distribution. Pakistan Journal of Statistics* 30(2): 273-284.

Jammalamadaka, S.R. & SenGupta, A. 2001. *Topics in Circular Statistics.* London: World Scientific.

Khanabsakdi, S. 1996. *Inferential statistics for concentration of directional data using the chisquare distribution. The Philippine Statistician* 44-45(1-8): 61-67.

Kucwaj, J.C., Reboul, S., Stienne, G., Choquel, J.B. & Benjelloun, M. 2017. Circular regression applied to GNSS-R phase altimetry. *Remote Sensing* 9(7): 651.

Letson, D. & McCullough, B.D. 1998. *Better confidence intervals: The double bootstrap with no pivot. American Journal of Agricultural Economics* 80(3): 552-559.

Loh, W.Y. 1991. *Bootstrap calibration for confidence interval construction and selection. Statistica Sinica* 1(2): 477-491.

Lv, X., Zhang, G. & Xu, X. 2017. *Bootstrap-calibrated empirical likelihood confidence intervals for the difference between two Gini indexes. Journal of Economic Inequality* 15(2): 195-216.

Mardia, K.V. 1972. *Statistics of Directional Data.* London: Academic Press.

Moslim, N.H., Zubairi, Y.Z., Hussin, A.G., Hassan, S.F. & Yunus, R.M. 2017. On the approximation of the concentration parameter for von Mises distribution. *Malaysian Journal of Fundamental and Applied Science* 13(4-1): 390-393.

Polo, M.E., Pozo, M. & Quiros, E. 2018. Circular statistics applied to the study of the solar radiation potential of rooftops in a medium-sized city. *Energies* 11(10): 2813.

Satari, S.Z., Zubairi, Y.Z., Hussin, A.G. & Hassan, S.F. 2015. Some statistical characteristic of Malaysian wind direction recorded at maximum wind speed: 1999-2008. *Sains Malaysiana* 44(10): 1521-1530.

Stephens, M.A. 1969. *A goodness-of-fit statistic for the circle, with some comparisons. Biometrika* 56(1): 161-168.

Stock, H., van Emmerik, R., Wilson, C. & Preatoni, E. 2018. Applying circular statistics can cause artefacts in the calculation of vector coding variability: A bivariate solution. *Gait & Posture* 65: 51-56.

Yan, L., Xiong, L.H., Liu, D.D., Hu, T.S. & Xu, C.Y. 2016. Frequency analysis of nonstationary annual maximum flood series using the time-varying two-component mixture distributions. *Hydrology Processes* 31(1): 69-89.

Nor Hafizah Moslim
Institute of Graduate Studies
Universiti Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Nor Hafizah Moslim
Faculty of Industrial Sciences & Technology
Universiti Malaysia Pahang
Lebuhraya Tun Razak
26300 Gambang, Pahang Darul Makmur
Malaysia

Yong Zulina Zubairi* & Siti Fatimah Hassan
Centre of Foundation Studies for Sciences
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Abdul Ghapor Hussin & Nurkhairany Amyra Mokhtar
Faculty of Defense Sciences and Technology
National Defense University of Malaysia
Kem Sungai Besi
57000 Kuala Lumpur, Federal Territory
Malaysia

*Corresponding author; email: yzulina@um.edu.my

Received: 23 September 2018
Accepted: 6 March 2019